

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Hötten, Gertrud  
Neichardt, Helge  
Paulista, Michael
- (ii) TITLE OF INVENTION: NEW GROWTH/DIFFERENTIATING FACTOR OF  
THE TGF- FAMILY
- (iii) NUMBER OF SEQUENCES: 40
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
(B) STREET: 655 Fifteenth Street N.W. Suite 330  
(C) CITY: Washington  
(D) STATE: D.C.  
(E) COUNTRY: U.S.A.  
(F) ZIP: 20005-5701
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: UNASSIGNED  
(B) FILING DATE: 31-AUG-1999  
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: 08/288,508  
(B) FILING DATE: 10-AUG-1994
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: DE P 43 26 829.3  
(B) FILING DATE: 10-AUG-1993
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: DE P 44 18 222.8  
(B) FILING DATE: 25-MAY-1994
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: DE P 44 20 157.5  
(B) FILING DATE: 09-JUN-1994
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: KITTS, Monica Chin  
(B) REGISTRATION NUMBER: 36,105  
(C) REFERENCE/DOCKET NUMBER: P564-9022
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (202)638-5000  
(B) TELEFAX: (202)638-4810

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2703 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCATGGCCTC GAAAGGGCAG CGGTGATTTT TTTCACATAA ATATATCGCA CTTAAATGAG 60  
 TTTAGACAGC ATGACATCAG AGAGTAATTA AATTGGTTTG GGTTGGAATT CCGTTTCCAA 120  
 TTCCTGAGTT CAGGTTTGTA AAAGATTTTT CTGAGCACCT GCAGGCCTGT GAGTGTGTGT 180  
 GTGTGTGTGT GTGTGTGTGT GTGTGTGTGA AGTATTTTCA CTGGAAAGGA TTCAAAATA 240  
 GGGGGAAAAA AAAACTGGAG CACACAGGCA GCATTACGCC ATTCTTCCTT CTTGGAAAAA 300  
 TCCCTCAGCC TTATACAAGC CTCCTTCAAG CCTCAGTCA GTTGTGCAGG AGAAAGGGGG 360  
 CGGTTGGCTT TCTCCTTTCA AGAACGAGTT ATTTTCAGCT GCTGACTGGA GACGGTGCAC 420  
 GTCTGGATAC GAGAGCATT CCACATAGGG ACTGGATACA AACACACACC CGGCAGACTT 480  
 CAAGAGTCTC AGACTGAGGA GAAAGCCTTT CCTTCTGCTG CTA CTGCTGC TGCCGCTGCT 540  
 TTTGAAAGTC CACTCCTTTC ATGGTTTTTC CTGCCAAGC AGAGGCACCT TTGCTGCTGC 600  
 CGCTGTTCTC TTTGGTGTCA TTCAGCGGCT GGCCAGAGGA TGAGACTCCC CAACTCCTC 660  
 ACTTTCTTGC TTTGGTACCT GGCTTGGCTG GACCTGGAAT TCATCTGCAC TGTGTTGGGT 720  
 GCCCCTGACT TGGGCCAGAG ACCCCAGGGG ACCAGGCCAG GATTGGCCAA AGCAGAGGCC 780  
 AAGGAGAGGC CCCCCCTGGC CCGGAACGTC TTCAGGCCAG GGGGTCACAG CTATGGTGGG 840  
 GGGGCCACCA ATGCCAATGC CAGGGCAAAG GGAGGCACCG GGCAGACAGG AGGCCTGACA 900  
 CAGCCCAAGA AGGATGAACC CAAAAGCTG CCCCCCAGAC CGGGCGGCCC TGAACCCAAG 960  
 CCAGGACACC CTCCCCAAAC AAGGCAGGCT ACAGCCCGGA CTGTGACCCC AAAAGGACAG 1020  
 CTTCCCGGAG GCAAGGCACC CCCAAAAGCA GGATCTGTCC GCAGCTCCTT CCTGCTGAAG 1080  
 AAGGCCAGGG AGCCCGGGCC CCCACGAGAG CCCAAGGAGC CGTTTCGCCC ACCCCCCATC 1140  
 ACACCCACG AGTACATGCT CTCGCTGTAC AGGACGCTGT CGGATGCTGA CAGAAAGGGA 1200  
 GGCAACAGCA GCGTGAAGTT GGAGGCTGGC CTGGCCAACA CCATCACCAG CTTTATTGAC 1260  
 AAAGGGCAAG ATGACCGAGG TCCCGTGGTC AGGAAGCAGA GGTACGTGTT TGACATTAGT 1320  
 GCCCTGGAGA AGGATGGGCT GCTGGGGGCC GAGCTGCGGA TCTTGCGGAA GAAGCCCTCG 1380  
 GACACGGCCA AGCCAGCGGC CCCCAGGAGC GGGCGGGCTG CCCAGCTGAA GCTGTCCAGC 1440

TGCCCCAGCG GCCGGCAGCC GGCCTCCTTG CTGGATGTGC GCTCCGTGCC AGGCCTGGAC 1500  
 GGATCTGGCT GGGAGGTGTT CGACATCTGG AAGCTCTTCC GAAACTTTAA GAACTCGGCC 1560  
 CAGCTGTGCC TGGAGCTGGA GGCCTGGGAA CGGGGCAGGG CCGTGGACCT CCGTGGCCTG 1620  
 GGCTTCGACC GCGCCGCCCG GCAGGTCCAC GAGAAGGCCC TGTTCCTGGT GTTTGGCCGC 1680  
 ACCAAGAAAC GGGACCTGTT CTTAATGAG ATTAAGGCCC GCTCTGGCCA GGACGATAAG 1740  
 ACCGTGTATG AGTACCTGTT CAGCCAGCGG CGAAAACGGC GGGCCCCACT GGCCACTCGC 1800  
 CAGGGCAAGC GACCCAGCAA GAACCTTAAG GCTCGCTGCA GTCGGAAGGC ACTGCATGTC 1860  
 AACTTCAAGG ACATGGGCTG GGACGACTGG ATCATCGCAC CCCTTGAGTA CGAGGCTTTC 1920  
 CACTGCGAGG GGCTGTGCGA GTTCCCATTG CGTCCACC TGGAGCCAC GAATCATGCA 1980  
 GTCATCCAGA CCCTGATGAA CTCCATGGAC CCGAGTCCA CACCACCCAC CTGCTGTGTG 2040  
 CCCACGCGGC TGAGTCCCAT CAGCATCCTC TTCATTGACT CTGCCAACAA CGTGGTGTAT 2100  
 AAGCAGTATG AGGACATGGT CGTGGAGTCG TGTGCTGCA GGTAGCAGCA CTGGCCCTCT 2160  
 GTCTTCCTGG GTGGCACATC CCAAGAGCCC CTCCTGCAC TCCTGGAATC ACAGAGGGGT 2220  
 CAGGAAGCTG TGGCAGGAGC ATCTACACAG CTTGGGTGAA AGGGGATTCC AATAAGCTTG 2280  
 CTCGCTCTCT GAGTGTGACT TGGGCTAAAG GCCCCCTTTT ATCCACAAGT TCCCCTGGCT 2340  
 GAGGATTGCT GCCCGTCTGC TGATGTGACC AGTGGCAGGC ACAGGTCCAG GGAGACAGAC 2400  
 TCTGAATGGG ACTGAGTCCC AGGAAACAGT GCTTTCCGAT GAGACTCAGC CCACCATTTC 2460  
 TCCTCACCTG GGCCTTCTCA GCCTCTGGAC TCTCTAAGC ACCTCTCAGG AGAGCCACAG 2520  
 GTGCCACTGC CTCCTCAAAT CACATTTGTG CCTGGTGAAT TCCTGTCCCT GGGACAGTTG 2580  
 AGAAGCTGAC TGGGCAAGAG TGGGAGAGAA GAGGAGAGGG CTTGGATAGA GTTGAGGAGT 2640  
 GTGAGGCTGT TAGACTGTTA GATTAAATG TATATTGATG AGATAAAAAG CAAAAGTGTG 2700  
 CCT 2703

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 501 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Leu Pro Lys Leu Leu Thr Phe Leu Leu Trp Tyr Leu Ala Trp  
 1 5 10 15

Leu Asp Leu Glu Phe Ile Cys Thr Val Leu Gly Ala Pro Asp Leu Gly

20 25 30  
 Gln Arg Pro Gln Gly Thr Arg Pro Gly Leu Ala Lys Ala Glu Ala Lys  
 35 40 45  
 Glu Arg Pro Pro Leu Ala Arg Asn Val Phe Arg Pro Gly Gly His Ser  
 50 55 60  
 Tyr Gly Gly Gly Ala Thr Asn Ala Asn Ala Arg Ala Lys Gly Gly Thr  
 65 70 75 80  
 Gly Gln Thr Gly Gly Leu Thr Gln Pro Lys Lys Asp Glu Pro Lys Lys  
 85 90 95  
 Leu Pro Pro Arg Pro Gly Gly Pro Glu Pro Lys Pro Gly His Pro Pro  
 100 105 110  
 Gln Thr Arg Gln Ala Thr Ala Arg Thr Val Thr Pro Lys Gly Gln Leu  
 115 120 125  
 Pro Gly Gly Lys Ala Pro Pro Lys Ala Gly Ser Val Pro Ser Ser Phe  
 130 135 140  
 Leu Leu Lys Lys Ala Arg Glu Pro Gly Pro Pro Arg Glu Pro Lys Glu  
 145 150 155 160  
 Pro Phe Arg Pro Pro Pro Ile Thr Pro His Glu Tyr Met Leu Ser Leu  
 165 170 175  
 Tyr Arg Thr Leu Ser Asp Ala Asp Arg Lys Gly Gly Asn Ser Ser Val  
 180 185 190  
 Lys Leu Glu Ala Gly Leu Ala Asn Thr Ile Thr Ser Phe Ile Asp Lys  
 195 200 205  
 Gly Gln Asp Asp Arg Gly Pro Val Val Arg Lys Gln Arg Tyr Val Phe  
 210 215 220  
 Asp Ile Ser Ala Leu Glu Lys Asp Gly Leu Leu Gly Ala Glu Leu Arg  
 225 230 235 240  
 Ile Leu Arg Lys Lys Pro Ser Asp Thr Ala Lys Pro Ala Ala Pro Gly  
 245 250 255  
 Gly Gly Arg Ala Ala Gln Leu Lys Leu Ser Ser Cys Pro Ser Gly Arg  
 260 265 270  
 Gln Pro Ala Ser Leu Leu Asp Val Arg Ser Val Pro Gly Leu Asp Gly  
 275 280 285  
 Ser Gly Trp Glu Val Phe Asp Ile Trp Lys Leu Phe Arg Asn Phe Lys  
 290 295 300  
 Asn Ser Ala Gln Leu Cys Leu Glu Leu Glu Ala Trp Glu Arg Gly Arg  
 305 310 315 320  
 Ala Val Asp Leu Arg Gly Leu Gly Phe Asp Arg Ala Ala Arg Gln Val  
 325 330 335

His Glu Lys Ala Leu Phe Leu Val Phe Gly Arg Thr Lys Lys Arg Asp  
 340 345 350  
 Leu Phe Phe Asn Glu Ile Lys Ala Arg Ser Gly Gln Asp Asp Lys Thr  
 355 360 365  
 Val Tyr Glu Tyr Leu Phe Ser Gln Arg Arg Lys Arg Arg Ala Pro Leu  
 370 375 380  
 Ala Thr Arg Gln Gly Lys Arg Pro Ser Lys Asn Leu Lys Ala Arg Cys  
 385 390 395 400  
 Ser Arg Lys Ala Leu His Val Asn Phe Lys Asp Met Gly Trp Asp Asp  
 405 410 415  
 Trp Ile Ile Ala Pro Leu Glu Tyr Glu Ala Phe His Cys Glu Gly Leu  
 420 425 430  
 Cys Glu Phe Pro Leu Arg Ser His Leu Glu Pro Thr Asn His Ala Val  
 435 440 445  
 Ile Gln Thr Leu Met Asn Ser Met Asp Pro Glu Ser Thr Pro Pro Thr  
 450 455 460  
 Cys Cys Val Pro Thr Arg Leu Ser Pro Ile Ser Ile Leu Phe Ile Asp  
 465 470 475 480  
 Ser Ala Asn Asn Val Val Tyr Lys Gln Tyr Glu Asp Met Val Val Glu  
 485 490 495  
 Ser Cys Gly Cys Arg  
 500

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGAATTCGCA TGCCATGGTC GACG

24

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CTTGAGTACG AGGCTTTCCA CTG

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTCGCATGC CATGGTCGAC GAAG

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGAGCCCACG AATCATGCAG TCA

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACAGCAGGTG GGTGGTGTGG ACT

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 44 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGAATTCGCA TGCCATGGTC GACGAAGCTT TTTTTTTTTT TTTT

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCAGCAGCCC ATCCTTCTCC

20

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCCAGGGCAC TAATGTCAAA CACG

24

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTAATGTCA AACACGTACC TCTG

24

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCGGCCGCT

10

## (2) INFORMATION FOR SEQ ID NO:13:





## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 60
- (C) OTHER INFORMATION: /note= "Xaa is a gap inserted only for comparison purposes"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Cys Lys Arg His Pro Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asn  
 1 5 10 15  
 Asp Trp Ile Val Ala Pro Pro Gly Tyr His Ala Phe Tyr Cys His Gly  
 20 25 30  
 Glu Cys Pro Phe Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala  
 35 40 45  
 Ile Val Gln Thr Leu Val Asn Ser Val Asn Ser Xaa Lys Ile Pro Lys  
 50 55 60  
 Ala Cys Cys Val Pro Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu  
 65 70 75 80  
 Asp Glu Asn Glu Lys Val Val Leu Lys Asn Tyr Gln Asp Met Val Val  
 85 90 95  
 Glu Gly Cys Gly Cys Arg  
 100

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 60
- (C) OTHER INFORMATION: /note= "Xaa is a gap inserted only for comparison purposes"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Cys Arg Arg His Ser Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asn  
 1 5 10 15  
 Asp Trp Ile Val Ala Pro Pro Gly Tyr Gln Ala Phe Tyr Cys His Gly  
 20 25 30  
 Asp Cys Pro Phe Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala  
 35 40 45  
 Ile Val Gln Thr Leu Val Asn Ser Val Asn Ser Xaa Ser Ile Pro Lys  
 50 55 60  
 Ala Cys Cys Val Pro Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu  
 65 70 75 80  
 Asp Glu Tyr Asp Lys Val Val Leu Lys Asn Tyr Gln Glu Met Val Val  
 85 90 95  
 Glu Gly Cys Gly Cys Arg  
 100

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 102 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Cys Lys Lys His Glu Leu Tyr Val Ser Phe Arg Asp Leu Gly Trp Gln  
 1 5 10 15  
 Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala Ala Phe Tyr Cys Asp Gly  
 20 25 30  
 Glu Cys Ser Phe Pro Leu Asn Ala His Met Asn Ala Thr Asn His Ala  
 35 40 45  
 Ile Val Gln Thr Leu Val His Leu Met Phe Pro Asp His Val Pro Lys  
 50 55 60  
 Pro Cys Cys Ala Pro Thr Lys Leu Asn Ala Ile Ser Val Leu Tyr Phe  
 65 70 75 80  
 Asp Asp Ser Ser Asn Val Ile Leu Lys Lys Tyr Arg Asn Met Val Val  
 85 90 95  
 Arg Ser Cys Gly Cys His  
 100

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 102 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys Arg Lys His Glu Leu Tyr Val Ser Phe Gln Asp Leu Gly Trp Gln  
1 5 10 15  
Asp Trp Ile Ile Ala Pro Lys Gly Tyr Ala Ala Asn Tyr Cys Asp Gly  
20 25 30  
Glu Cys Ser Phe Pro Leu Asn Ala His Met Asn Ala Thr Asn His Ala  
35 40 45  
Ile Val Gln Thr Leu Val His Leu Met Asn Pro Glu Tyr Val Pro Lys  
50 55 60  
Pro Cys Cys Ala Pro Thr Lys Leu Asn Ala Ile Ser Val Leu Tyr Phe  
65 70 75 80  
Asp Asp Asn Ser Asn Val Ile Leu Lys Lys Tyr Arg Asn Met Val Val  
85 90 95  
Arg Ala Cys Gly Cys His  
100

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 102 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Cys Lys Lys His Glu Leu Tyr Val Ser Phe Arg Asp Leu Gly Trp Gln  
1 5 10 15  
Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala Ala Tyr Tyr Cys Glu Gly  
20 25 30  
Glu Cys Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr Asn His Ala  
35 40 45  
Ile Val Gln Thr Leu Val His Phe Ile Asn Pro Glu Thr Val Pro Lys  
50 55 60  
Pro Cys Cys Ala Pro Thr Gln Leu Asn Ala Ile Ser Val Leu Tyr Phe  
65 70 75 80  
Asp Asp Ser Ser Asn Val Ile Leu Lys Lys Tyr Arg Asn Met Val Val  
85 90 95  
Arg Ala Cys Gly Cys His  
100

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATGAATTCCC ATGGACCTGG GCTGGMAGKA MTGGAT

36

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ACGTGGGGTG GAATGACTGG AT

22

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATATTGGCTG GAGTGAATGG AT

22

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATGTGGGCTG GAATGACTGG AT

22

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

22

ACCTGGGCTG GCAGGACTGG AT

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

22

AGGACCTCGG CTGGAAGTGG AT

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

22

GGGATCTAGG GTGGAAATGG AT

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

22

AGGATCTGGG CTGGAAGTGG GT

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

22

AGCTGGGCTG GGAACGGTGG AT

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

22

ACATCGGCTG GAATGACTGG AT

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

22

TCATCGGCTG GAACGACTGG AT

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

29

ATGAATTCGA GCTGCGTSGG SRCACAGCA

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

21

GAGTTCTGTC GGGACACAGC A

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

21

CATCTTTTCT GGTACACAGC A

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

21

CAGTTCAGTG GGCACACAAC A

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

21

GAGCTGCGTG GGCGCACAGC A

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

21

CAGCGCCTGC GGCACGCAGC A

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TAAATCTTGG GACACGCAGC A

21

## (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAGGTCCTGG GGCACGCAGC A

21

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCCTGGGAGA GCAGCACAGC A

21

## (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CAGCTTGGTG GGCACACAGC A

21

## (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:

"GAGCTTG"



- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CAGCTTGGTG GGAATGCAGC A

667663" 031130200

SEQ ID NO.1

TYPE OF SEQUENCE: nucleic acid sequence

NAME AND ORIGIN: MP-52 DNA

LENGTH: 2703 nucleotides

CCATGGCCTCGAAAGGGCAGCGGTGATTTTTTTCACATAAATATATCGCACTTAAATGAG  
TTTAGACAGCATGACATCAGAGAGTAAATTAATTTGGTTTGGGTTGGAATTCGGTTTCCAA  
TTCCCTGAGTTTCAGGTTTGTAAAAGATTTTCTGAGCACCTGCAGGCCTGTGAGTGTGTGT  
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGAAGTATTTTCACTGGAAAGGATTCAAACTA  
GGGGGAAAAAAACTGGAGCACACAGGCAGCATTAAGCCATTCTTCCTTCTTGGAAAAA  
TCCCTCAGCCTTATACAAGCCTCCTTCAAGCCCTCAGTCAGTTGTGCAGGAGAAAGGGGG  
CGGTTGGCTTTCTCCTTTCAAGAACGAGTTATTTTCAGCTGCTGACTGGAGACGGTGCAC  
GTCTGATACGAGAGCATTTCCACTATGGGACTGGATACAAACACACACCCGGCAGACTT  
CAAGAGTCTCAGACTGAGGAGAAAGCCTTTCTTCTGCTGCTACTGCTGCTGCCGCTGCT  
TTTGAAAGTCCACTCCTTTTCATGGTTTTTTCCTGCCAAACCAGAGGCACCTTTGCTGCTGC  
CGCTGTTCTCTTTGGTGTGATTTCAGCGGCTGSCCAGAGGATGAGACTCCCCAAACTCCTC  
ACTTTCTGTGCTTTGGTACCTGGCTTGGCTGGACCTGGAATTCATCTGCACTGTGTGGGT  
GCCCCTGACTTGGGCGCAGAGACCCAGGGGACCTAGGCCAGGATTGGCCAAAGCAGAGGCC  
AAGGAGAGGCCCCCTGGCCCCGGAACGTCTTCAGGCCAGGGGTCACAGCTATGGTGGG  
GGGGCCACCAATGCCAATGCCAGGGCAAAGGGAGGCACCGGGCAGACAGGAGGCCTGACA  
CAGCCCCAAGAAGGATGAACCCAAAAGCTGCCCCCAGACCGGGCGGCCCTGAACCCAAAG  
CCAGGACACCTCCCCAAACAAGGCAGGCTACAGCCCGGACTGTGACCCCAAAGGACAG  
CTTCCCCGAGGCAAGGCACCCCCAAAAGCAGGATCTGTCCCCAGCTCCTTCTGCTGAAG  
AAGGCCAGGGAGCCCCGGGCCCCCAGAGAGCCCAAGCAGCCGTTTCGCCACCCCCCATC  
ACACCCACAGAGTACATGCTCTCGCTGTACAGGACGCTGTCCGATGCTGACAGAAAGGGA  
GGCAACAGCAGCGTGAAGTTGGAGGCTGGCCTGGCCAAACCATCACCAAGCTTTATTGAC  
AAAGGGCAAGATGACCGAGGTCCCGTGGTCAGGAAGCAGAGGTACGTGTTTGACATTAGT  
GCCCTGGAGAAGGATGGGCTGCTGGGGGCCGAGCTGCGGATCTTGCGGAAGAAGCCCTCG  
GACACGGCCAAGCCAGCGSCCCCCGAGGCGGGCGGGCTCCCCAGCTGAAGCTGTCCAGC  
TGCCCCAGCGGCCGGCAGCGGCCCTCCTTGCTGGATGTGCGCTCCGTGCCAGGCCTGGAC  
GGATCTGGCTGGGAGGTGTGACATCTGGAAGCTCTTCCGAAACTTTAAGAACTCGGCC  
CAGCTGTGCCCTGGAGCTGGAGGCTGGGAACGGGGCAGGGCCTGGACCTCCGTGGCCTG  
GGCTTCGACCGCGCGCCCCGGCAGGTCCACGAGAAGGCCCTGTTCCTGGTGTTTGGCCGC  
ACCAAGAAACGGGACCTGTCTTTAATGAGATTAAAGGCCCGCTCTGGCCAGGACGATAAG

036450.05495263

2  
C13

ACCGTGTATGAGTACCTGTTTCAAGCAGCGGCGAAAACGGGCGGGCCCCACTGGCCACTGGC  
CAGGGCAAGCGACCCAGCAAGAACCCTTAAGGCTGGCTGCAGTGGGAAGGCACTGCATGTC  
AACTTCAAGGACATGGGCTGGGACGACTGGATCATCGCACCCCTTGAGTACGAGGCTTTC  
CACTGCGAGGGGCTGTGCGAGTTCCCATTTGCGCTCCCACTGGAGCCCAAGATCATGCA  
GTTCATCCAGACCCCTGATGAACCTCATGGACCCCGAGTCCACACCACCCACCTGCTGTGTG  
CCCACGCGGCTGAGTCCCATCAGCATCTCTTCATTGACTCTGCCAACAACGTGGTGTAT  
AAGCAGTATGAGGACATGGTCTGTGGAGTCTGTGTGGCTGCAGGTAGCAGCACTGGCCCTCT  
GTCTTCCTGGGTGGCACATCCCAAGAGCCCCCTTCCTGCACTCCTGGAATCAAGAGGGGT  
CAGGAAGCTGTGGCAGGAGCATCTACACAGCTTGGGTGAAAGGGGATTCCAATAAGCTTG  
CTCGCTCTCTGAGTGTGACTTGGGCTAAAGGCCCCCTTTTATCCACAAGTTCCCCCTGGCT  
GAGGATGTCTGCCCGTCTGCTGATGTGACCACTGGCAGGCACAGGTCCAGGGAGACAGAC  
TCTGAATGGGACTGAGTCCCAGGAACAGTGTCTTCCGATGAGACTCAGCCCACCATTTTC  
TCTCACCTGGGCCTTCTCAGCCTCTGGACTCTCCTAAGCACCTCTCAGGAGAGCCACAG  
GTGCCACTGCTCCTCAAAATCACATTTGTGCTGGTACTTCTGTCCCTGGGACAGTTG  
AGAAGCTGACTGGGCAAGAGTGGGAGAGAAGAGGAGAGGGCTTGGATAGAGTTGAGGAGT  
GTGAGGCTGTAGACTGTAGATTAAATGTATATGATGAGATAAAAAGCAAAACTGTG  
CCT

0030450.003450

SEQ ID NO:2

TYPE OF SEQUENCE: amino acid sequence

NAME AND ORIGIN: MP-52 protein

LENGTH: 501 amino acids

*Seq  
C13  
included*

MRLPKLLTFL LWYLAULDLE FICTVLGAPD LGQRPQGTRP GLAKAEAKER  
PPLARNVFRP GGHSYGGGAT NANARAKGGT GQTGGLTQPK KDEPKKLPPR  
PGGPEPKPGH PPQTRQATAR TVTPKGQLPG GKAPPKAGSV PSSFLLKKAR  
EPGPPREPKE PFRPPPIITPH EYMLSLYRTL SDADRKGGS SVKLEAGLAN  
TITSFIDKGQ DDRGPVVRKQ RYVFDISALE KDGLLGAE LR ILRKKPSDTA  
KPAAPGGGRA AQLKLSSCS GRQPASLLDV RSVPGLDGSG WEVFDIWKLF  
RNFKNQAQLC LELEAWERGR AVDLRGLGFD RAARQVHEKA LFLVFGRTKK  
RDLFFNEIKA RSGQDDKTVY EYLFSQRRKR RAPLATRQ GK RPSKNLKARC  
SRKALHVNFK DMGWDDWIIA PLEYEAFHCE GLCEFPLRSH LEPTNHAVIQ  
TLMNSMDPES TPPTCCVPTR LSPISILFID SANNVVYKQY EDMVVESCGC R

---